1019-1002

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of: Fultheim, Shai	Application No.: 10/828,465
Filed: 04/21/04	Art Unit: 2128
For: Cluster Based Operating System-Agnostic Virtual Computing System	Examiner: David Silver

DECLARATION UNDER 37 CFR 1.132

I, the undersigned, Boaz Yehuda, hereby declare as follows:

1) I am making this Declaration in support of the patentability of the claims in U.S. Patent Application 10/828,465 (referred to hereinafter as "the Application"). Specifically, this Declaration will set forth my opinion, based on my own first-hand experience, that the invention defined by the claims in the Application and embodied in the vSMP product sold by ScaleMP (the assignee of the Application), answers a real market need that could not be satisfied by prior art solutions.

2) I am not an employee of ScaleMP and have no economic interest in the company. I have not received compensation for my services in preparing this Declaration.

3) I have worked in the computer industry for 23 years, specializing in sales and service of server systems. For six years I worked for Sun Microsystems as Sun Israel Country Manager, leading its business and sales in Israel. During this period, I

became familiar with ScaleMP. My detailed *curriculum vitae* is attached hereto as Exhibit A.

4) Prior to ScaleMP, all the multi-processor servers (consisting of more than one physical processor) I had seen were made with custom electronics, either on a single proprietary PCB board (motherboard) or with special proprietary interconnects between such custom motherboards. Although these systems were capable of providing high computing power - greater than commodity PC computers/servers - their high cost made them impractical for all but the most high-end applications.

5) I was initially skeptical when I first heard about ScaleMP's vSMP technology. I had never before seen or heard of a solution in which commodity servers and commodity interconnects could be used to create multi-processor servers via use of software - be it virtual machine-type software or any other class. Although such a solution would have helped us in serving end-user needs in high-power computing applications, the idea itself struck me as a nearly impossible goal to achieve for software technology. I was therefore favorably surprised when I first saw vSMP working.

6) In my experience, vSMP technology has succeeded in meeting the long-felt need in the IT market for high-power, low-cost multiprocessor systems. The technology is not only novel, but is also highly attractive for end-users and system manufacturers in the IT market:

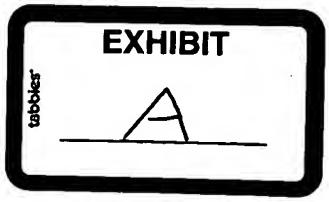
- It reduces time to market for superior commodity processor technology to find its way into multi-processor servers.

- It dramatically lowers the costs of designing, creating, procuring, and deploying multi-processor servers.
- It enables the use of existing software assets, without requiring intrusive changes such as recompilation and operating system modifications.

7) I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and conjecture are thought to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application of any patent issued thereon.

Boaz Yehuda, Citizen of Israel
16 Jerusalem Street, Kfar Saba, Israel

Date: 17/10/09



Boaz Yehuda
16 Jerusalem st.
Kfar Saba
Israel

Tel . (home): +972-9-7673868
Tel. Mobile: +972-54-617778

CURRICULUM VITA

Professional Profile:

Top experience of 20 plus years in management sales, marketing and service in the Computers and Storage industry.
Interpersonal and team leadership abilities.
Negotiation experience.
Strong Presentation skills.
Record of achievements in penetrating new markets.
Record achievements in general management
Proven record of working with complex accounts.
Good understanding of the Israeli market.
International sales management experience.
Well recognized and established relationships in Israeli IT industry.
Well recognized for managing complex operations.

Professional Experience:

2009- To date **BizMe2: Chairman**

BizMe2 is One Stop shop that gather all the information and recommendations about convention and tradeshows that occur all over the world, it offers its social internet network platform for visitors that want to meet other colleagues with the same interest, it helps them to schedule appointments and enlarge their business opportunity. Moreover it is centralizing all the booking services such as (flight tickets, hotels, transportation, and tradeshow registration) at one place.

2008- To date **Liyha B.A Business development: CEO**

Liyha has two business strictures:
a. Working with Israeli companies helping them establish international presence through a network of partners.
b. Development of web software as service for physiotherapists

2002- 2008 **Sun Microsystems: Sun Israel General Manager**

Responsible for managing Sun Microsystems activities in Israel. With more than 150 employees in Israel, divided for development center, sales and services, was responsible for development of the new software product penetration strategy, which resulted winning 100% of the ISP market. During that period was responsible for

doubling the high-end server install base and renewing major OEM contracts through deeper penetration to those customers. Deployed complex organizational changes that allowed Sun Israel to align with the new corporate strategy. During that period Sun Israel was outstanding for its innovation, tools and new process that led to major growth in profitability(more than doubling the contribution) and revenue (doubling the revenue). Sun Israel under my leadership won it's largest ever deals.

1999-2002 **Sun Microsystems: Product Sales and Marketing Manager Mediterranean Region**

Responsible for building and managing the product sales and marketing team.
Responsible for the product sales strategies and deep involvement in all major deals and product events in the region.
Responsible for product training in the region.
Responsible for product marketing communications in the region.
Managing successfully several major crises in the region. For example: Solving the Partner Orange crises that generated over $3M sales with high end servers and services for Sun Israel during FY02.
Developing the Joint Account Team methodology for account management. Method used by Sun Israel and Sun Turkey to manage their most important accounts.

1998-1999 **Sun Microsystems: Storage Business Development Manager Mediterranean Region.**

Responsible for driving storage sales and strategies in the region. During that period, storage sales in the region had more than doubled, surpassing the sales goals by 80%. Attach rate was close to 30%.
Was rewarded for the best product sales driver in SEAME for FY99.

1996-1998: **EMC Israel: Open Systems Sales Manager**

Responsible for sales, strategies, partners and channels for Open Systems in EMC Israel. During that time EMC increased sales in the open systems market from $100K to $4M. Successfully penetrated many large accounts like IAI, Paz, Partner-Orange and other telecommunications operators. Personally developed the application based sales approach in EMC Israel, winning most of ERP and Data-warehouse sales bids.

1995 - 1996: **E&M Computing Ltd: Vertical Sales Manager**

Managing a group of sales persons responsible for various sectors of the Israeli market including defense, utilities and other scientific and technical areas at E&M computing LTD., sole distributor of SUN Microsystems in Israel at that time.

1992 - 1995: **E&M Computing Ltd: Sales Engineer**

Full time Sales Engineer to various sectors in the Israeli market. In 1995, named distinguished sales engineer of SUN for the whole Mediterranean region. Responsible for more than 95% of Sun server sales in region during that period.

1989 - 1992: **E&M Computing Ltd.: Maintenance Associate Engineer**

While studying at the Technion, Israel Institute of Technology in Haifa, Responsible for service of SUN computers in northern Israel. During the summer vacations, managed the central maintenance lab at E&M Computing.

1986 - 1989: **E&M Computing Ltd.: Maintenance Associate Engineer .**

Responsible for service of SUN workstations and servers. Responsible for developing a hardware maintenance section at E&M Computing Ltd.

Education:

1988 - 1992:	**B.Sc.:, Technion, Israel Institute of Technology, Haifa.** Electrical & Electronics Engineering
1985 - 1987:	**Tel Aviv University school for Associate Engineering.** Associate Electronics Engineer (with distinction).

Military Service:

1981 - 1985:	Military service that included variety of professional people command rules. Work that consisted multitasking and crucial decision making under time pressure. The last year of service served as instructor and a course commander.
Languages:	Excellent Hebrew, excellent English and good Arabic.

EXHIBIT

tabbies

B

HP and ScaleMP put you in charge

All-in-one "Shorty" with ScaleMP vSMP Foundation

A supercomputer-in-a-box workgroup solution makes HPC simple and personal.

Affordable HPC at your command

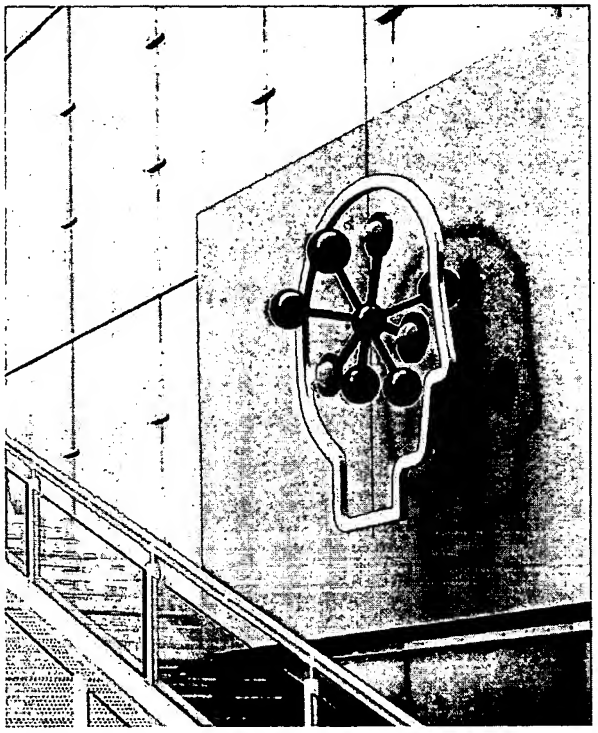
A High Performance Computing (HPC) capability is essential for organizations to compete in a demanding global economy. Clustering has helped make HPC more affordable and accessible for parallel workloads, but remains complex. In addition, solutions for workloads requiring large memory have been too expensive until now.

HPC data center capability, outside the data center

HP and ScaleMP extend the affordability and accessibility of clusters to large memory workloads, while simplifying administration. ScaleMP vSMP Foundation seamlessly aggregates multiple industry-standard x86 systems into a single virtual system. The combination of vSMP Foundation and the HP Cluster Platform Workgroup system ("Shorty") enables workgroups and small organizations to deploy high performance capability outside the data center, running mixed large memory and parallel workloads. The "Shorty" utilizes HP BladeSystems in a small, compact design that conserves space while decreasing power consumption and heat generation.

Solution benefits

- Simple and rapid deployment on a compact, powerful system via Cluster Platform Express
- Run parallel applications on up to 128 cores
- Run extremely large jobs and models with up to 1 TB of shared memory
- Single system simplicity
- Cost-effective, power-efficient and flexible HP BladeSystem design



Working together with HP

HPC solutions enable rapid advancements innovation, cost-efficiency and productivity. Using proven HPC methodologies, organizations of all sizes can speed and optimize their engineering, financial, drug discovery and development processes to drive greater success. Working collaboratively with our partners, we apply our vast knowledge and experience in HPC to deliver complete solutions that help you tap into the strength and flexibility of supercomputing powered by highly reliable HP servers.

ScaleMP™ | hp®

Single virtual HPC system

vSMP Foundation, deployed on the HP Cluster Platform Workgroup system, offers tremendous price/performance advantages for the HPC market. In essence, it provides a unique way to leverage entry-level systems to reduce the total cost of ownership (TCO). It delivers the operational simplicity of traditional shared-memory systems while keeping the acquisition cost associated with clusters. With an ability to support both parallel and large memory workloads, the solution is a supercomputer-in-a-box. The HP Cluster Platform Workgroup system, leveraging the top selling HP BladeSystem technology, requires no special power, cooling or staff and can be deployed outside the data center. With a footprint of less than two square feet, it can deliver almost a TFLOP of power, with up to eight nodes using HP ProLiant BL460c Servers, all communicating over the integrated InfiniBand network incorporated into the workgroup system.

HP products supported

Our solutions are built on best-selling HP ProLiant servers, world-renowned for reliability and availability. From server blades to clustered servers, HP ProLiant servers provide the utmost confidence for your business. With new Intel® Xeon™ based HP ProLiant server models, HP further extends the advantages of x86 computing—delivering more cost-effective, industry-standard solutions for applications requiring expanded memory and outstanding price/performance. HP ProLiant servers are your best choice for long-term dependability, flexibility and growth.

Operating systems supported

Our Linux-based solutions are built on innovative software and standards-based servers—delivered by service professionals with extensive experience. By working with HP, we can leverage its worldwide leadership in Linux to provide optimized solutions based on HP Open Source Integrated Portfolio (OSIP) and HP Open Source Middleware Stacks (OSMS). You can trust our combined expertise to provide the proven, cost-effective Linux solutions you need to drive a fast return on your IT investments.

Service and support

ScaleMP solutions build on HP BladeSystem, the foundation for HP Cluster Platform Workgroup Systems, preconfigured, turn-key HPC clusters available from HP and our partners. ScaleMP installation services and support are available directly from ScaleMP and its partners.

Building on the value of strong relationships

By working collaboratively with HP, you can leverage extensive resources, deep experience and broad industry knowledge to provide innovative solutions that drive positive business results and long-term value. With a proven record of success in virtually every industry in every region worldwide, HP understands what you need to increase your organization's success. Together, we can deliver best-fit technology and services to meet your unique business requirements.

ScaleMP is a leader in virtualization for high-end computing, providing higher performance and lower TCO. Using software to replace expensive and proprietary symmetric multiprocessing systems (SMP), ScaleMP offers a new computing paradigm, aggregating industry-standard systems into a single virtual x86 system to tackle the most difficult workloads.

Working together to drive success

By combining standards-based and high-performance technologies from HP with our leading-edge applications and comprehensive services, we can deliver a powerful, flexible solution that meets your technology requirements. With a tailor-made solution designed by business partners you trust, you can face your toughest business challenges—including increasing productivity, quickening time to market and reducing operating costs—through greater efficiency, rapid discovery and reduced development cycles.

Technology for better business outcomes

To learn more, visit www.hp.com
www.scalemp.com

4AA2-3301ENW, November 2008

ScaleMP™ | hp®

EXHIBIT

tabbies

C

SIMPLY PERFORM!

TIRED OF WAITING FOR SHARED DATACENTER RESOURCES?

TIRED OF WAITING FOR YOUR SIMULATIONS TO COMPLETE?

WOULD YOU PREFER PROGRAMMING WITH OPENMP?

WANT TO INSTALL A SUPERCOMPUTER IN LESS THAN A DAY?

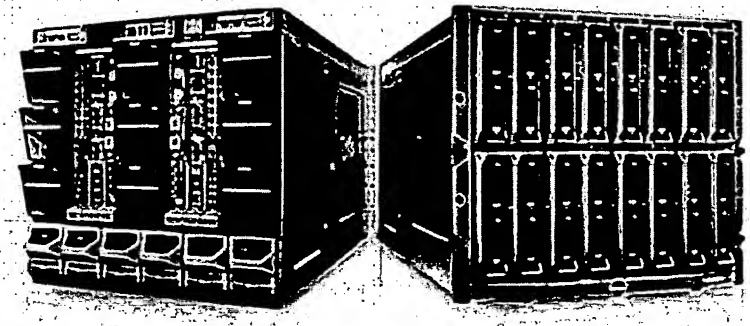
DREAMING ABOUT INFINIBAND SPEED WITHOUT HEADACHES?

WANT TO SIMPLIFY CLUSTER STORAGE BY A FACTOR OF 16?

Dell PowerEdge M1000e

Single Virtual System

Optimized for High Performance Technical Computing (HPTC)



1.6 TFLOPS
+ 128 Cores
+ 3TB Shared Memory
+ 19TB Internal Storage

1 Operating System

DELL

Dell PowerEdge M1000e Single Virtual System At A Glance

The Dell PowerEdge M1000e Single Virtual System is an x86 supercomputer with up to 32 processors (128 cores) and up to 3TB of shared memory running a single instance of a Linux operating system. Based on ScaleMP's vSMP Foundation™ software that creates a single virtual system by aggregating multiple Dell MxxO blade servers within a M1000e chassis, this system is ideally suited for high performance computing applications in the financial services, life sciences, engineering and educational institutions. It offers significant price/performance advantage, power consumption savings, and higher density over traditional and proprietary SMP systems. Finally, high performance SMP's are affordable again...

PROBLEMS SOLVED

Reducing cost of traditional SMP Systems

Traditional high-end x86 systems with four, eight or more sockets are based on lengthy and proprietary R&D developments that must be passed on to end users. These systems also usually incorporate older generation components and chip speeds. This results in expensive solutions that have lower compute density, consume more power, and cost more on a per-socket basis compared to dual-socket systems. ScaleMP's vSMP Foundation, by aggregating the more powerful dual-socket servers into a single virtual system offers better performance, scalability, yet at a lower cost!

Reducing complexity of cluster deployments

Today's clusters are designed to provide high-density coupled with excellent performance and power efficiency. However, management costs are high: multiple Operating Systems required, replication of applications and content. In addition, a complex high-speed cluster file-systems or proprietary external storage solutions must be implemented. Applications are limited to the memory footprint per system. ScaleMP's vSMP Foundation converts clusters into affordable SMPs: single Operating System, internal storage and large memory. Same components – just simpler to manage and run.

BENEFITS

Large memory

The Dell PowerEdge M1000e Single Virtual System enables an application to use the aggregated memory of all the blades in the system. In the extreme, a single application process can leverage up to 3TB RAM. Large memory also reduces the need to use external high-performance storage systems for swap or scratch space. Application runtime is dramatically reduced by running simulations with in-core-solvers or by using memory instead of swap for large models. In addition, both traditional SMP codes (OpenMP) and distributed applications (MPI) run at optimal performance on the same physical infrastructure.

Compute & memory intensive applications

For workloads that require a high core count coupled with shared memory, users have traditionally acquired proprietary shared-memory systems. The PowerEdge M1000e Single Virtual System provides a very cost effective x86 alternative to these expensive RISC systems. As opposed to traditional SMP or NUMA architecture where memory bandwidth decreases as the machine scales, it combines memory-bandwidth across boards and demonstrates close to linear memory bandwidth scaling. The scalable memory bandwidth delivers excellent performance for memory intensive applications.

Ease of use

For workloads that otherwise require a scale-out approach, the Dell PowerEdge M1000e Single Virtual System provides ease of use by having a single system to manage, compared the complexities involved with managing a cluster. A single system removes the need for cluster file systems, cluster interconnect issues, application provisioning and installation and update of multiple operating systems and applications. This results in significant savings at installation, and ongoing operations.

Simplified I/O architecture

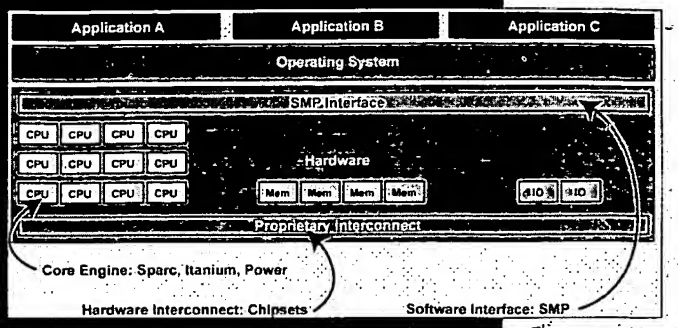
High-bandwidth I/O requirements in a scale-out model can be complex and costly, usually involving HBA's, and FC switch infrastructure. The Dell PowerEdge M1000e Single Virtual System aggregates each individual server's network and storage interfaces. I/O resource consolidation reduces the number of drivers, HBA's, NIC's, cables, and switch ports and all the associated maintenance overhead. The user needs fewer I/O devices to purchase, manage and service.

Improved Utilization

For large compute farms deployments the Dell PowerEdge M1000e Single Virtual System becomes an attractive alternative for organizations that need to run hundreds or thousands of simulations at once. As opposed to hundreds of servers, where each server operates at 80 percent utilization (to allow for runtime peaks), fewer larger systems can run more applications on the same footprint.

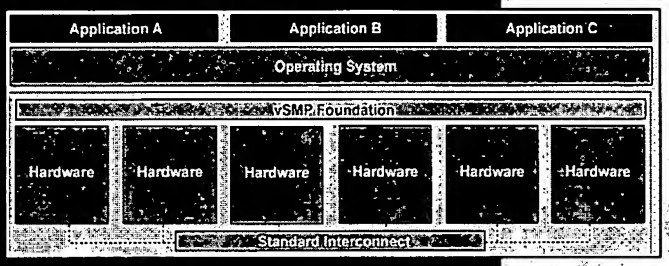
How Does It Work?

Traditional SMP systems run a single operating system (OS) which interacts with the system using a well-defined hardware interface. This interface provides the OS with predefined services to use and control the hardware, including hardware detection and probing, memory ordering semantics, I/O space access and interrupt delivery mechanisms. An example of such hardware interface is the Intel's Multi-Processor Specification (MP Spec) which defines a standard interface between the hardware and the OS to make it easy for the OSVs and OEMs to quickly support a wide range of platforms with one OS version, a benefit they already enjoy in the Uni-processor desktop market for Intel Architecture CPUs. In essence, the MP Spec brings the same "shrinkwrap" benefits of the desktop market to the MP market. For a traditional SMP system, such interface is implemented in a silicon chipset. In addition to the hardware interface, an SMP system consists of CPUs, memory and I/O subsystems, all connected together with a proprietary backplane or interconnect such as Intel's FSB (Front Side Bus), AMD's HT (Hyper-Transport), SUN's CrossBar SGI's NUMALINK and IBM's XA. The proprietary backplane (system interconnect) is where today's SMP systems differ the most from one another.



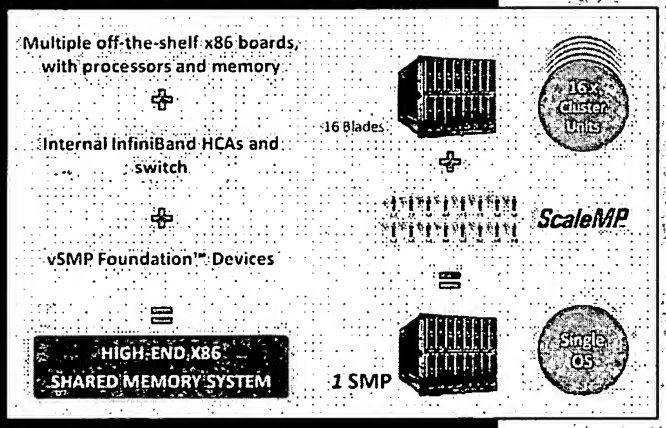
THE SCALEMP VERSATILE SMP™ (VSMP) ARCHITECTURE

The vSMP architecture utilizes off-the-shelf components and does not require any custom parts. Its key value is the utilization of software to provide the chipset services that are otherwise required in creating traditional multi-processor systems. vSMP Foundation provides cache coherency, shared I/O and the system interfaces (BIOS, ACPI) , which are required by the OS. *The vSMP architecture is implemented in a completely transparent manner*; no additional device drivers are required and no modifications to the OS or the applications are necessary.



From a hardware perspective, vSMP Foundation requires:

- Multiple Dell x86 systems or blades
- InfiniBand HCA's, cables and switch to interconnect the systems or blades
- vSMP Foundation Devices – Flash-based storage devices (one per board/system) with the appropriate vSMP Foundation product supporting the specific Dell products



From a system perspective, vSMP Foundation provides:

- One single system: once loaded in memory of each system boards, vSMP Foundation aggregates all the resources of the multiple physical systems, initializes the interconnect fabric, and creates the required BIOS and ACPI environment to provide the OS a coherent image of a single virtual system. vSMP Foundation then uses a software-interception engine in the form of a Virtual Machine Monitor (VMM) to provide a uniform execution environment.

- Coherent Memory: vSMP Foundation maintains cache coherency between the individual boards using multiple advanced coherency algorithms that operate concurrently on a per-block basis, based on real-time memory activity access patterns. vSMP Foundation leverages board local-memory together with best-of-breed caching algorithms to offset the effect of interconnect latencies.

- Shared I/O: vSMP Foundation aggregates I/O resources across all boards into a unified PCI hierarchy and presents them as a common pool of I/O resources to the OS and applications

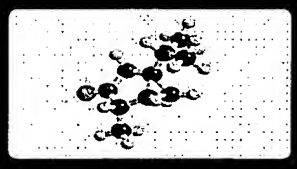
Frequently Asked Questions

Question	Answer
What node is the master node?	There is no concept of a master node in shared memory systems (as there is only sa ingle node). vSMP Foundation however has a concept of primary device (inserted in the first board which contains the configuration information, and is where the system keyboard, video and mouse should be used) and secondary devices (for all other boards).
Where does the operating system run?	The operating system runs over the entire system. It can boot from local drives or from the network.
Do applications or operating systems need to be changed or modified to run on the system?	The systems run standard Linux operating systems distributions. Any x86 binary runs as-is, exactly as if it were running on the standard dual socket x86 server. However, as vSMP Foundation enables systems with up to 128 processor cores and terabytes of RAM, several tuning and optimization should be considered to reap maximum performance. ScaleMP offers application execution guidelines for significant number of applications as well as tuning suggestion for the Linux kernel.
Can I disable vSMP Foundation to run MPI?	vSMP Foundation can be certainly be disabled if one wants to run native cluster. However, as MPI applications run at the same performance level with, or without vSMP Foundation, most users just run MPI applications in shared memory.
How stable is this technology?	vSMP Foundation first appeared on the market in 2005, and is now implemented in production at over 100 sites worldwide. End-users include manufacturing companies, higher education institutions, life-sciences and pharmaceutical organizations as well as leading financial institutions.
What is the latency for off-board memory access?	The ScaleMP versatile SMP (vSMP) architecture is hybrid COMA-NUMA architecture. As such there is no notion of fixed latency for memory access. In essence, the vSMP architecture minimizes the number of times a processor fetches memory from a remote board. Think of it as additional, board-level, cache.

Benefits By Industry Sectors

Manufacturing

Leverage high memory bandwidth and large number of cores for structural and impact analysis, and computational fluid dynamics applications. Take advantage of large shared memory for implicit analysis, pre-processing and post-processing.



Energy

Most reservoir and volume interpretation applications require large memory and high memory bandwidth. Seismic processing requires a large number of processors. The vSMP architecture is ideal for such applications.

Numerical simulations

Flexibility to run numerical simulations using all the memory in the system, multiple processes in parallel sharing the memory, or one or more jobs running in multi-processors mode.

Higher education and research

Dynamic adjustment to the mix of multidisciplinary applications, as well as ever-changing research priorities; from jobs that require a large memory footprint, high number of processors, or small to medium simulations in throughput mode.

Life sciences

Flexible, high-performance system to run a large number of disparate legacy, OpenMP, MPI applications in one system, by leveraging high number of processors, large memory or bandwidth or a combination thereof.

Electronic Design Automation (EDA)

Utilize the same hardware infrastructure for large shared memory processing during validation phases (prior to tape-out) and running multiple concurrent user jobs on large core count in day-to-day use.

Bottom Line: Get A Headstart On The Competition

End user	Rubber Manufacturing Corp	Engineering Services Company	Formula 1 Team
Existing infrastructure	High-core count Itanium system	Multiple 2-socket workstations	Large Memory Itanium system
Challenges	Need to run thread-based applications, such as Gaussian, as well as MPI-based Computational Chemistry applications	• Existing models (Abaqus) grow fast and no longer fit engineers' workstations • Need to run large simulations in batch at night • No in-house skills to run x86 InfiniBand cluster and cannot afford RISC systems	• Need to generate large meshes as part of pre-processing of whole car simulation (FLUENT TGrid) • Mesh requirements are over 200 GB in size • Wants to standardize on x86 architecture due to lower cost
Solution	• 8 Xeon Quad-core processors • 32-cores • 128 GB RAM	• 8 Xeon Dual-core processors • 16 cores • 128GB RAM	• 24 Xeon Processors • 96 cores • 384 GB RAM
Performance Benefits	• Significantly faster than the existing IA64 SMP • Performance is comparable to cluster performance with similar-hardware	• Significantly faster than existing workstations • Performance is comparable to cluster performance	Evaluated and proven to be faster than alternative systems (x86 and non-x86)
Operational Benefits	No IT resources required for day-to-day operation	No IT required for day-to-day operation	Similar to existing system
Capital Expenditure Benefits	Significant savings compared to upgrade or replacement of existing system with IA64		Significant savings compared to existing and alternative systems considered
Versatility		Interactive jobs during the day, batch jobs at night	Used for large memory jobs (TGrid) and regular FLUENT (MPI) solvers on large number of cores
Investment protection		Expected to double the resources	Headroom for growth

Available Configurations

Specifications

- Min. / Max. boards: 2 / 16
- Min. / Max. memory (GB) per board: 4 / 256
- Min. / Max. processors per board: 1 / 4
- Min. / Max. cores per board: 1 / 16
- Max. system memory: 4 TB
- Max. system processors: 64
- Max. system cores: 128

Supported platforms: http://www.ScaleMP.com/spec

Screaming Performance

Facts

- The fastest x86 system by memory bandwidth
- The 15th fastest system by memory bandwidth globally
- The fastest x86 system by SPEC CPU2006
- The largest shared-memory x86 system by RAM size and core count

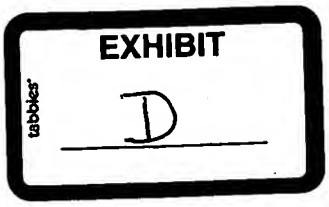
For More Information

Dell

- EMEA: Paul Brook (Paul_Brook@Dell.com) EMEA HPC Programme Manager
- USA: Karl Cain (Karl_Cain@Dell.com) HPC Business Development Manager

ScaleMP

- dell@scalemp.com



CRAY
THE SUPERCOMPUTER COMPANY

News Release

Cray and ScaleMP Announce Strategic Alliance

SEATTLE, WA and CUPERTINO, CA, Mar 05, 2009 (MARKET WIRE via COMTEX) -- Global supercomputer leader Cray Inc. (NASDAQ: CRAY) and ScaleMP, a leading provider of virtualization solutions for high-end computing, today announced a strategic alliance to offer joint solutions based on the Cray CX1(TM) deskside supercomputer and ScaleMP's vSMP Foundation. Available immediately, the joint solution will target the High Performance Computing (HPC) segment allowing customers to operate a shared-memory, deskside supercomputer that scales up to 128 cores and 1TB of shared memory.

The Cray and ScaleMP strategic alliance is focused on enabling supercomputing at the workstation level. The combined Cray CX1 system and vSMP Foundation solution enables workgroups and small organizations to deploy high performance computing capabilities that harness the power of multiple processors while simplifying their operational environment. This solution is versatile, able to run a variety of Linux(R) workloads such as large memory, parallel workloads and high core count shared memory applications, and delivers excellent performance across many programming models ranging from MPI, OpenMP and legacy code.

"Cray and ScaleMP are addressing important requirements for HPC by offering a personal supercomputer workstation," said Earl Joseph, IDC Program Vice President. "Many departmental and work group users of HPC applications have been constrained by the lack of in-house skills to move up to clusters from workstations. This solution will allow these users to more easily scale up their simulations and models and boost productivity and competitiveness without the added complexity."

"I am very excited about our collaboration with Cray," said Shai Fultheim, founder and CEO of ScaleMP. "Cray is synonymous with excellence in the high-end supercomputer segment. This announcement enables HPC customers to get Cray performance at the deskside, in a cost-effective, workstation-like simplicity. Cray customers will be leveraging the capabilities of vSMP Foundation to achieve a flexible compute resource capable of solving bigger problems -- accelerating time to market and innovation."

"The ScaleMP vSMP Foundation virtualization software is an excellent fit for the Cray CX1, which was designed specifically to harness HPC for individuals and departmental workgroups," said Ian Miller, senior vice president of the productivity solutions group and marketing at Cray. "By creating a single shared memory virtual system, the joint solution can now support large memory and large core count workloads in addition to parallel workloads, while simplifying the installation and management of the system."

vSMP Foundation aggregates multiple industry-standard, off-the-shelf x86 servers (rack mounted or blade systems) into one single virtual high-end system for the HPC market. vSMP Foundation provides customers with an alternative to traditional expensive symmetrical multiprocessor (SMP) systems and also offers simplified clustering infrastructure with a single operating system. It currently allows customers to create a single virtual SMP system with up to 32 sockets (128 cores) and up to 4 TB of shared memory in an energy-efficient, dense package.

About Cray Inc.

As a global leader in supercomputing, Cray provides highly advanced supercomputers and world-class services and support to government, industry and academia. Cray technology enables scientists and engineers to achieve remarkable breakthroughs by accelerating performance, improving efficiency and extending the capabilities of their most demanding applications. Cray's Adaptive Supercomputing vision will result in innovative next-generation products that integrate diverse processing technologies into a unified architecture, allowing customers to surpass today's limitations and meeting the market's continued demand for realized performance. Go to www.cray.com for more information.

Cray CX1 Supercomputer

The Cray CX1 product is an affordably-priced, deskside supercomputer. Easy to configure, deploy, administer and use, it is the "right size" in performance, functionality and cost for a wide range of users, from the single user who wants a personal supercomputer to a department of users as a shared clustered resource. Equipped with powerful Intel Xeon(R) processors and Windows(R) HPC Server 2008 or Red Hat Enterprise Linux with Clustercorp Rocks+, the Cray CX1 product offers performance leadership across a broad range of applications

and standard benchmarks. For organizations wanting to harness HPC without the complexity of traditional clusters, the Cray CX1 supercomputer delivers the power of a high performance cluster with the ease-of-use and seamless integration of a workstation.

About ScaleMP

ScaleMP is the leader in virtualization for high-end computing, providing maximum performance and lower Total Cost of Ownership (TCO). The innovative Versatile SMP(TM) (vSMP) architecture aggregates multiple x86 systems into a single virtual x86 system, delivering an industry-standard, high-end symmetric multiprocessor (SMP) computer. Using software to replace custom hardware and components, ScaleMP offers a new, revolutionary computing paradigm. The company is backed by Sequoia Capital, Lightspeed Venture Partners, TL Ventures, and ABS Ventures. For more information, please call +1 (408) 342-0330 or visit www.ScaleMP.com.

Cray Media:
Nick Davis
206/701-2123
nickd@cray.com

ScaleMP-Media:
Amar Rao
408/342-0330
PR@ScaleMp.com

SOURCE: Cray Inc.

mailto:nickd@cray.com mailto:PR@ScaleMp.com

Next Generation Data Center Environment for HPC

Enabling the Dynamic Data Center

3LEAFSYSTEMS

Copyright

3Leaf Systems, the 3Leaf Systems logo, 3Leaf Systems software, and Virtual Compute Environment are trademarks or registered trademarks of 3Leaf Systems, Inc. in the United States and/or other countries.

All other brands, products or service names are or may be registered trademarks, trademarks, or service marks of their respective owners and are used to identify the products or services of their respective owners.

Notice: This document is for informational purpose only and does not set forth any warranty, express or implied, concerning any equipment, equipment feature, or service offered or to be offered by 3Leaf Systems. 3Leaf Systems reserves the right to make changes to this document at any time, without notice, and assumes no responsibility for its use. This informational document describes features that may not be currently available. Contact 3Leaf Systems for information on feature and product availability.

Export of technical data contained in this document may require an export license from the United States Government.

3Leaf Systems

3255-1 Scott Blvd., Suite 200

Santa Clara, CA 95054

Phone: 408.572.5900

Fax: 408.727.2008

www.3leafsystems.com

Virtualization and HPC

HPC solutions today face challenges in flexibility, cost, software development time and performance. In the enterprise world, some of these very same constraints have been solved by vendors with software virtualization solutions (e.g., VMware or Xen). However, the HPC market is different from the traditional enterprise market in that higher utilization of computing resources, a big problem for enterprise customers, is a distant second priority to performance. As such, virtualization solutions popular in the enterprise data center have little to offer to the HPC user, whose primary concern is performance or possibly performance per dollar or performance per watt.

Indeed, software virtualization solutions such as hypervisors attempt to multiplex multiple virtual machines onto a single physical machine to improve the physical machine's utilization. However, in the HPC environment, often a single job can fully utilize the machine. For many HPC applications, the problem is not machine utilization.

In addition, the focus on performance, performance per dollar, performance per watt, and flexibility has precipitated the trend towards cluster based computing. Cluster based computing takes advantage of inexpensive, commodity computing and network resources (e.g. x86 computers and high speed LANs) to form large pools of interconnected compute resources. Because a cluster can be sized to fit the application, these clusters have significant cost and flexibility advantages over the big iron supercomputers they replace.

Moving towards cluster based computing also means shifting to a new memory model. The memory model provided by many traditional super computers is coherent shared memory. The memory model presented by HPC clusters is distributed memory on independent nodes, connected by a low latency message passing network. The mapping of the traditional shared memory paradigm to that of distributed memory and message passing is not without difficulty. Some problems do not partition easily into computational threads that have low communication overhead. In others, while the partitioning may be obvious, the amount of communication or the latency presented by communication becomes a performance bottleneck.

Besides mapping the communication paradigm from shared memory to message passing, a programmer needs to deal with the hard resource constraints of the individual compute nodes when writing or porting an application to an HPC cluster. The compute nodes in traditional cluster computers have hard limits on the number of CPU cores available, the amount of memory and the amount of I/O.

3Leaf's Dynamic Data Center

3Leaf's Dynamic Data Center ("DDC") enables efficient scale-up computing across multiple commodity servers. This greatly expands the flexibility and scalability, enabling multiple physical servers to be dynamically grouped together into a single logical server. By allowing complete flexibility in how CPU and memory resources are provisioned, both capital and operational expenditures are significantly reduced.

3Leaf's unique technology for enabling CPU and memory virtualization includes both ASIC (the 3Leaf Coherent NIC) and software technology (the 3Leaf Distributed Machine Monitor). The 3Leaf Coherent NIC is the first of its kind, and extends the coherency domain of an x86 processor across multiple x86 commodity platforms using commodity switch fabrics. By doing this, the physical boundaries of a server are expanded, and it provides the backbone for the virtualization of compute and memory resources to efficiently span multiple x86 physical servers. In other words, for the first time ever, a single Guest OS can span across many servers to utilize CPU or memory resources to handle spikes in application traffic.

In order to fully address the x86 market with its CPU and memory virtualization technology, 3Leaf has licensed the processor interconnects from both AMD and Intel. The AMD processor interconnect is Coherent HyperTransport (cHT), and Intel's is QuickPath Interconnect (QPI). Licensing these core pieces of technology enables 3Leaf to support both AMD and Intel based solutions, and support 100% of the x86 commodity server market.

Specific to HPC, 3Leaf's DDC removes hardware barriers of existing HPC clusters, enabling:

- construction of coherent memory domains across an HPC cluster,
- partitioning of memory resources into multiple coherency domains, unrestricted by physical constraints (e.g. server blade), and
- partitioning of CPU resources across a cluster, allowing for the best allocation of CPU resources to fit the application.

3Leaf's ability to fully virtualize all components of an x86 server into pools that can span across multiple physical machines is truly game changing and sets 3Leaf apart from the competition. The Dynamic Data Center enables 3Leaf to provide a complete server virtualization solution for commodity platforms, allowing compute, memory, and I/O resources to be dynamically allocated and de-allocated as required.

3Leaf's Network Shared Memory

3Leaf's Dynamic Data Center also brings powerful shared memory concepts to the cluster computing environment. Nodes within a cluster can collectively create a coherent shared memory region (i.e., network shared memory). Every node within the cluster can contribute memory to the network shared memory area, which can be directly addressed by every other node that is part of the network shared memory cluster.

The result is that there can be up to one terabyte of shared memory between specific configurations ranging from 2 to 256 nodes in the cluster. This memory is directly addressable within the network cluster and allows all the operations that shared memory is capable of, enabling much more scalable and lower-overhead data sharing and communication methods than can be expected in a traditional cluster environment.

3Leaf network shared memory offers support for read-modify-write operations at the hardware level to shared memory addresses. This ability is very useful in a cluster environment where synchronization is much more costly, typically being carried out through expensive distributed locking methods.

3Leaf's Dynamic Data Center for HPC

3Leaf's view of virtualization offers new opportunities for cluster based computers to provide the advantages of traditional shared memory supercomputers while leveraging the cost and scalability advantages of commodity based cluster computing. 3Leaf's DDC allocates compute and memory resources as required by the job. Unlike traditional virtualization solutions, 3Leaf's solution does not multiplex physical resources amongst competing virtual machines.

The 3Leaf DDC allows cluster CPU resources to be allocated to virtual servers, without regard to the physical nodes on which the CPUs reside. Servers can be created that match the HPC application's requirements, instead of HPC applications having to conform to the physical machine's topology. In other words, the machine conforms to the problem, the problem does not have to conform to the machine.

By providing coherent shared memory across the nodes in the cluster, the 3Leaf solution offers the performance of traditional shared memory with the capabilities and configuration flexibility of a cluster based computer. In addition, applications that use message passing can benefit from a shared memory implementation of the message passing library (e.g. OpenMPI with shared memory messaging) without code changes. Indeed, a shared memory implementation of message passing need not incur the network stack and scheduling overhead of distributed memory message passing solutions.

Therefore, 3Leaf network shared memory drastically modifies the paradigm for applications designed for cluster computing. With network shared memory, the cluster applications are freed from the overhead and latencies of message passing and distributed synchronization. The support for hardware read-modify-write operations in network shared memory enables very scalable and non-blocking synchronization methods formerly available only in traditional shared memory platforms. Furthermore, the memory devoted to the cluster computing environment in this way is utilized much more effectively. There is no longer the need to have multiple copies of the same data, along with the associated messaging overhead to replicate the data

between nodes. The lack of memory duplication leads to increasing a cluster's flexibility while reducing its cost.

Development Flow for HPC

The raw runtime for an HPC application only takes into account part of the solution cost. The development of HPC applications on cluster computers also needs careful consideration. For some HPC applications, development time is a significant time and cost component of the HPC problem. Broadly speaking, there are two paradigms to consider - parallel and pipelined.

In the parallel model, a computation problem is broken into identical or similar, computational threads. The data are then divided among the computing threads. These problems are often referred to as embarrassingly parallelizable, but there are also problems of this kind that have a high communication to computation ratio and do not run well on message passing clusters. An example of this kind of application is the simulation of fluid flows.

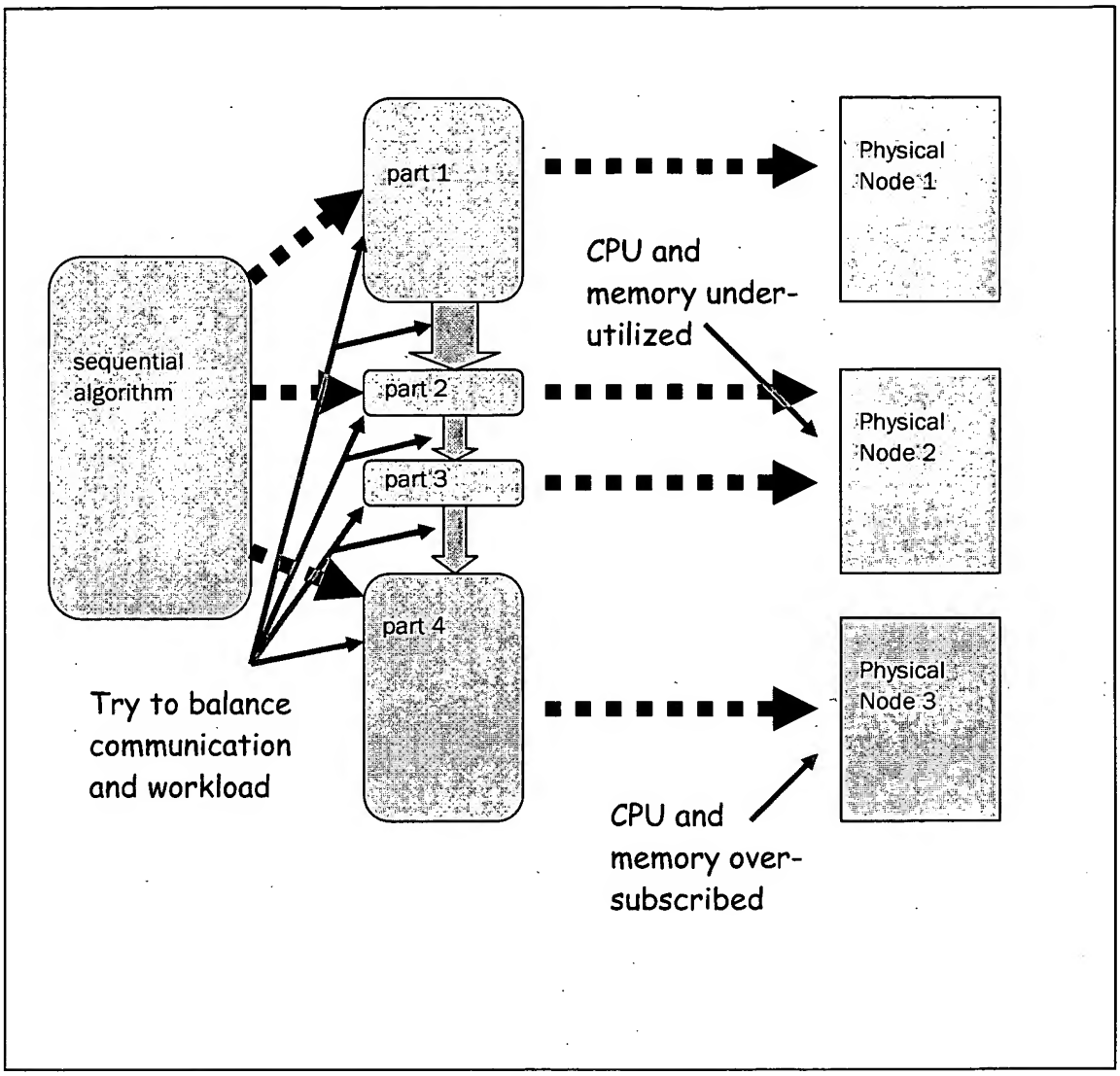
Another method of parallelizing an application is to divide the task into sequential subtasks that can map onto a computation pipeline. An example application that maps onto a pipeline nicely is image processing, where a series of image transformations can be chained together in a sequential pipeline.

The challenge for the HPC programmer is to map the problem into pipeline stages that satisfy three constraints:

1. minimize communication between stages
2. map the number of pipeline stages to match the number of computational nodes available
3. size each of the pipeline stages to fit within available CPU and memory resources of each node (optimally use all the available resources of each node)

3Leaf's DDC removes these constraints from the programmer by virtualizing CPU and memory resources on a cluster based computer. ***This allows programmers to spend more time focusing on solving the problem instead of mapping the problem onto a given machine configuration.*** Different CPU and memory configurations can be created to try different problem partitions, enabling application optimization without regard to the physical hardware constraints.

Figure 1 Mapping a Sequential Problem onto a Compute Pipeline



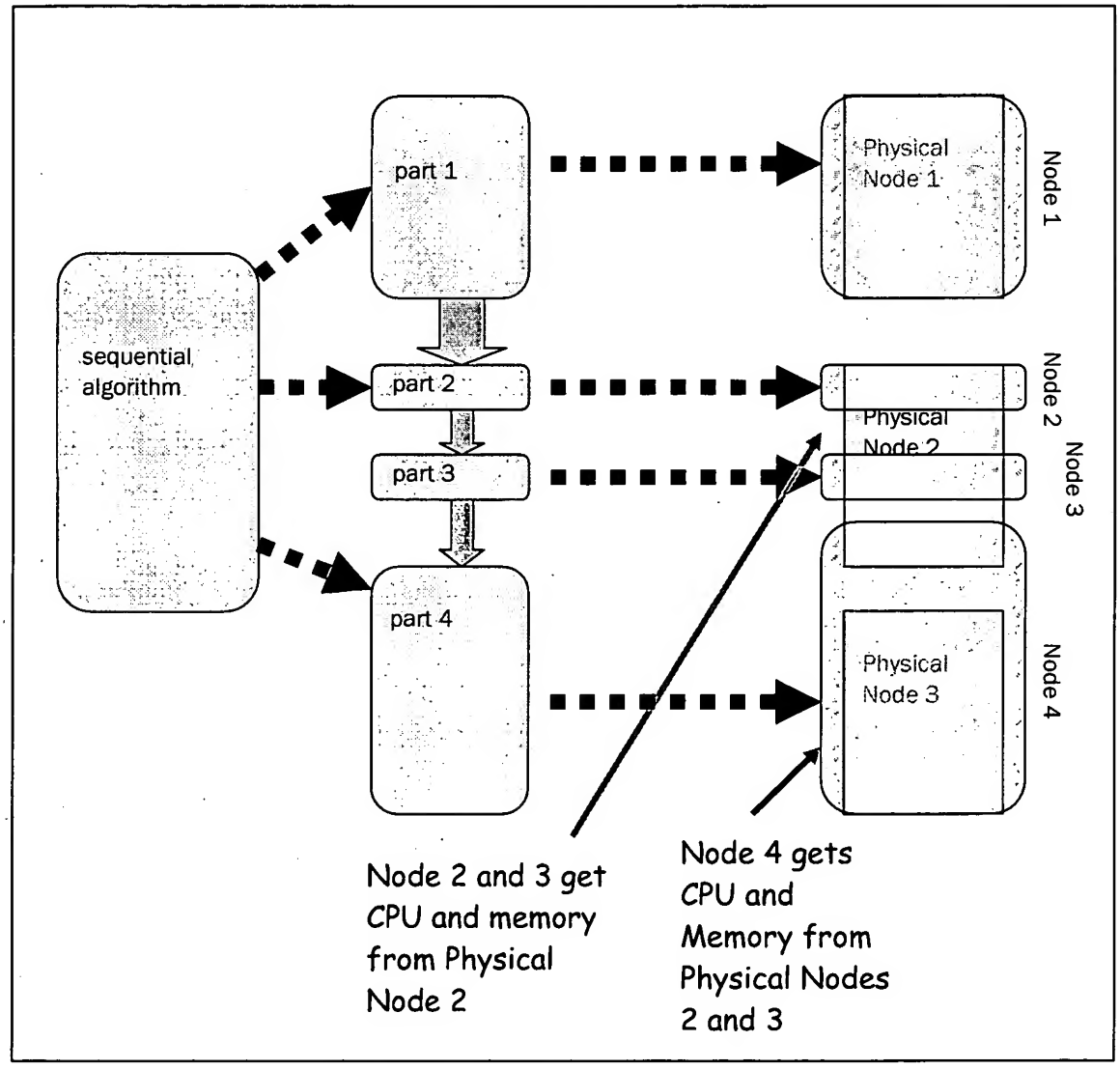
In a 3Leaf DDC, communication bandwidth and latency are reduced by allowing memory to be shared between nodes. The multi-node coherent memory allows for shared memory implementations of message passing (through a library like OpenMPI) or through shared memory exposed to the application. Thus, programmers can consider partitions that would have been impractical on traditional HPC clusters due to bandwidth and latency limitations of non-shared memory message passing.

Since 3Leaf implements Virtual Compute Nodes, the cores on individual servers can be grouped in a way that matches the number of compute nodes required by the application pipeline. For example, consider a physical machine consisting of 3 physical servers. If the problem better matches a machine that has 4 compute nodes, the CPUs on the three individual servers can be grouped into 4 logical servers. Indeed, one of the prime advantages of cluster computing over Scale-Up computing is that the cluster can be sized to match the problem. The 3Leaf Dynamic Data Center takes this notion even further, by allowing hardware clusters to be partitioned and allocated to fit each and every application.

Traditional HPC clusters have hard constraints on the number of CPU cores and amount of physical memory available. Each server has only a fixed amount of memory and CPU core resources. This forces the programmer to consider balancing the CPU and memory resources of each stage of the pipeline with available hardware resources. With the 3Leaf Dynamic Data Center, each computational node can be sized with the number of compute and amount of memory resources required.

Finally, the constraints on memory in traditional HPC clusters may lead some users or IT professionals to provision each node with the maximum amount of memory so that the programmer's burden is mitigated and system flexibility is increased. This solution results in wasted resources and results in higher capital and operating expenses. The 3Leaf Dynamic Data Center network shared memory solves this problem with its cluster wide shared memory, resulting in capital and operating expense efficiencies.

Figure 2 - 3Leaf Relieves Hardware Partition Limitations



Building on Industry Trends

Multiple industry trends have contributed to enable the 3Leaf solution for compute and memory virtualization:

Advanced commodity 64 bit processors

Today's processors include features that meet the needs of enterprise class data-centers, with seamless support for virtualization, 64 bit addressing, memory fault detection and recovery, hardware enforced secure portioning, multi-core going from 4 today to 6 and 8 tomorrow, and rapidly increasing cache sizes.

Network switches delivering increasing bandwidth with decreasing latency

As silicon process technology evolves, latencies for commodity switch chips continue to fall. Not only has the bandwidth of single chip switches reached Terabits per second, the time it takes from a signal arriving at the input pin to appear at the output pin is approaching the same order of magnitude as DRAM access times, with IB switches approaching switching times of 100nS and Ethernet switches in hot pursuit.

Enterprise Capable commodity operating systems

Commodity operating systems such as Windows and Linux now provide ccNUMA optimizations, and also support the hot plug and hot unplug of devices, CPU and memory. In addition, both Operating Systems and today's enterprise applications are fully compatible with virtualized servers (via a hypervisor)

3leaf builds on these trends

The Coherent NIC from 3Leaf lets threads running on cores in one server interact directly with threads and memory located on another server, and is enabled by the bandwidth and latency of modern switches. The Distributed Machine Monitor from 3Leaf insulates an Operating System from a distributed and changing set of core and memory resources, and is enabled by the advanced features of today's processors and enterprise class Windows and Linux operating systems.

Conclusion

HPC computing continues its migration towards cluster based solutions for flexibility, cost and performance reasons. However, traditional clusters do not provide the rich and critical performance features of coherent shared memory. 3Leaf's Dynamic Data Center brings the advantages of shared memory supercomputers to cluster based computers.

3Leaf is enabling the Dynamic Data Center with next generation server virtualization that addresses today's changing business needs by providing the on-demand resources and flexibility that can literally revolutionize operational and development efficiencies. Virtualization of CPU, memory, and I/O resources enables the creation of a pool of server resources that can span across multiple physical machines and be allocated or de-allocated as needed. Coherent shared memory among commodity x86 compute nodes provides a high performance compute cluster for HPC applications. Easy repurposing and migration allows machines to remain fully utilized as workloads change during the course of the day, week, or month. Fast and flexible provisioning allows machines to be mass deployed within a heterogeneous environment, drastically reducing the time and cost of developing applications and provisioning new servers and the applications they support.

3Leaf's Dynamic Data Center is truly dynamic, nimbly responding to the changing requirements of HPC applications, while drastically reducing both capital and operating expenses.